Multi-tissue proteomics identifies molecular signatures for sporadic and genetically defined Alzheimer disease cases

Carlos Cruchaga (ccruchaga@wustl.edu)
Washington University, School of Medicine https://orcid.org/0000-0002-0276-2899
Yun Ju Sung
Washington University in St. Louis
Chengran Yang
Washington University in St. Louis https://orcid.org/0000-0003-4577-5590
Fengxian Wang
Washington University in St. Louis https://orcid.org/0000-0002-7677-5406
Adam Suhy
Washington University, School of Medicine https://orcid.org/0000-0003-1282-6339
Joanne Norton
Washington University in St. Louis
Brenna Novotny
Washington University in St. Louis https://orcid.org/0000-0003-4732-1736
Abdallah Etelleb
Washington University in St. Louis
Anne Fagan
Washington University
Randall Bateman
Washington University in St. Louis https://orcid.org/0000-0002-7729-1702
Richard Perrin
Washington University in St. Louis https://orcid.org/0000-0002-3443-7716
John Morris
Washington University
Martin Farlow
Jasmeer Chhatwal
Harvard Med School
Helena Chui
University of Southern California
Herve Rhinn
Alector, Inc.

Celeste Karch
Washington University https://orcid.org/0000-0002-6854-5547
Suzanne Schindler
Washington University in St. Louis
Oscar Harari
Washington University School of Medicine, Saint Louis https://orcid.org/0000-0002-2635-6975

Biological Sciences - Article

Keywords: proteomic signatures, TREM2 variant carriers, sporadic AD

Posted Date: October 21st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-923492/v1

Abstract

Alzheimer disease (AD) is a heterogeneous disease with many genes are associated with AD risk. Most proteomic studies, while instrumental in identifying AD pathways and genes, focus on single tissues and sporadic AD cases. Multi-tissue proteomic signatures for sporadic and genetically defined AD (e.g., pathogenic variant carriers in APP and PSEN1/2 and risk variant carriers in TREM2) will illuminate the biology of this heterogeneous disease.1,2 Here, we present one of the largest multi-tissue proteomic profiles, accessible through our web portal, based on 1,305 proteins in brain (n=360), cerebrospinal fluid (CSF; n=717), and plasma (n=490) from the Knight Alzheimer Disease Research Center (Knight ADRC) and Dominantly Inherited Alzheimer Network (DIAN) cohorts.3-5 We identified proteomic signatures in brain, CSF, and plasma for sporadic AD status and replicated these findings in multiple, independent datasets. The area under the curve (AUC) for CSF proteins was 0.89 in discovery and 0.90 in the replication dataset, which was significantly higher than the AUC for CSF p-tau181/Aβ42 (AUC = 0.81; P = 2.4×10-6). We also identified a specific proteomic signature for TREM2 variant carriers that differentiated TREM2 variant carriers from sporadic AD cases and controls with high sensitivity and specificity (AUC = 0.81 - 1). In addition, the proteins that showed differential levels in sporadic AD were also altered in autosomal dominant AD, but with greater effect size (1.4 times, P = 3.8×10-5), and proteins associated with autosomal dominant AD, in brain tissue also replicated on CSF (p=1.36×10-9). Enrichment analyses highlighted several pathways including AD (calcineurin, APOE, GRN), Parkinson disease (α-synuclein, LRRK2), and innate immune response (SHC1, MAPK3, SPP1) for the sporadic AD or TREM2 variant carriers. Our findings show the power of multi-tissue proteomics' contribution to the understanding of AD biology and to the creation of tissue-specific prediction models for individuals with specific genetic profiles, ultimately supporting its utility in creating individualized disease risk evaluation and treatment.

Main Text

Alzheimer disease (AD), the most common cause of dementia, is a heterogeneous neurodegenerative disease characterized by neuronal loss, neuroinflammation, and memory decline.[6] Clinical sequelae reduce quality of life and cost the healthcare system up to $244 billion annually in the US, with additional impact on caregivers' emotional distress.[7] Proteomic studies have been instrumental in identifying biomarkers and pathways implicated in AD, but most have been limited to single tissues and only differentiate between sporadic AD cases and controls. Deep molecular characterization of controls, sporadic AD cases and genetically defined AD subtypes, such as individuals carrying pathogenic mutations in amyloid-beta precursor protein (*APP*) and presenilin genes (*PSEN1* and *PSEN2*) or high-effect risk variants in triggering receptor expressed on myeloid cells 2 (*TREM2*),[1,2,8] is critical for fully understanding the biology of this heterogeneous disease and for identifying novel molecular biomarkers and therapeutic targets. Here, we present the results of multi-tissue, high-throughput deep proteomic profiling. We identified proteins associated with AD status that replicated in external datasets. We not only identified proteomic profiles for sporadic AD, but also for individuals with AD-risk variants in *TREM2* and pathogenic variants in *APP* and *PSEN1/2*. These proteomic profiles enabled the creation of tissue-

specific prediction models and the identification of causal proteins and pathways for sporadic and genetically defined AD subtypes.

Study Design

To elucidate the downstream effects of genes and the functional mechanisms associated with AD, we generated high-throughput, deep proteomic profiles using SOMAscan targeting 1,305 proteins in brain tissue, cerebrospinal fluid (CSF), and plasma (Fig. 1).[5] These neurologically relevant tissues were obtained from well-characterized individuals with comprehensive clinical information about AD pathology and cognition in the Knight ADRC[3] and DIAN.[3,4,9,10] After stringent quality control (QC) and data cleaning, a total of 1,092 proteins from 360 brain tissues remained. These brain proteomic data include 24 individuals carrying autosomal dominant AD (ADAD) mutations in *APP* and *PSEN1/2*, 290 individuals with autopsy-confirmed AD, 21 *TREM2* variant carriers, and 25 cognitively normal individuals with no significant brain pathology (Table 1). CSF data contained 713 proteins from 176 individuals with a clinical diagnosis of AD, 47 *TREM2* variant carriers, and 494 cognitively normal individuals. Plasma data contain 931 proteins from 105 individuals with a clinical diagnosis of AD, 131 *TREM2* variant carriers, and 254 cognitively normal individuals (Fig. 1).

AD status was defined based on neuropathological examination for those samples with brain autopsy and clinical examination for those with CSF and plasma tissue. In this study we identified proteins with different levels in clinical AD cases vs. controls and not based on biomarker levels or the ATN framework, which combines the amyloid-β pathway (A), tau-mediated pathophysiology (T), and neurodegeneration (N), because one of the goals of this study was to compare the performance of the prediction models generated in this study with these well-accepted and validated CSF biomarkers (Aβ and p-tau181).

To validate and replicate proteins that were associated with AD, *TREM2* risk variant carriers or ADAD mutation carrier status, we followed two approaches: first, for sporadic AD and *TREM2* risk variant carriers, we identified the common set of proteins dysregulated in the three tissues (brain, CSF, and plasma). For ADAD, only high-throughput proteomic screening was performed on brain tissue. Those proteins that were associated with ADAD status in brain were analyzed in CSF from 289 ADAD mutation carriers and 184 non carriers from the DIAN study. Second, for sporadic AD, we used seven publicly available datasets to replicate our findings (Supplementary Table 1). For brain, we downloaded the mass-spectrometry data for following 6 studies: the Adult Changes In Thought (ACT), Banner Sun Health Research Institute (BANNER), Baltimore longitudinal study of aging (BLSA), Mayo Clinic (MAYO), Mount Sinai Brain Bank (MSBB), the Religious Orders Study and the Memory and Aging Project (ROSMAP). We

then performed differential abundance analysis jointly for a total of 10,078 proteins measured in 415 AD patients and 194 controls, called hereafter MassSpec Joint. For CSF, we obtained and analyzed Alzheimer's Disease Neuroimaging Initiative (ADNI) multiple reaction monitoring (MRM) proteomic data containing 320 proteins in 263 samples. We also used results based on BioFinder OLINK data from Whelan et al.[11] and Emory-ADRC mass-spectrometry data from Higginbotham et al.[12] For plasma, we downloaded and performed differential analysis on the AddNeuroMed SOMAscan 1.1K proteomic data that was processed and deposited by Sattlenecker et al.[13] We were not able to use public datasets to replicate the proteins dysregulated in *TREM2* or ADAD mutation carriers because there were not enough carriers in public datasets. Finally, we used the replicated proteins to generate prediction models and run pathway analyses. We combined the results from this study with our recent pQTL, colocalization, and Mendelian randomization findings to identify causal proteins.[5]

Multi-tissue proteomic signatures of AD

Sporadic AD cases

To identify multi-tissue proteomic signatures for clinical AD, we performed differential analysis with a subgroup of sporadic AD patients and healthy individuals in each of the three tissues, independently. Specifically, we performed a surrogate variable analysis (SVA)[14] to remove batch effects and other unmeasured heterogeneity in all three proteomic datasets. We then performed regression analysis of log-transformed protein abundance levels as a dependent variable and sporadic AD status as an independent variable while considering age, sex, and SVA as covariates.

Brain proteomic profiles for sporadic AD

In the brain, 12 proteins showed significant association for AD status after Bonferroni correction (Fig. 2a, Supplementary Table 2). We chose the Bonferroni-corrected threshold as it is more conservative than false discovery rate (FDR). All 12 proteins were nominally significant ($P < 0.05$) with other AD-related phenotypes including age at onset and AD neuropathology characteristics such as Braak scores and CDR at death (Supplementary Table 2, Supplementary Fig. 1-2). As we had proteomic data from CSF and plasma, we determined which proteins are also associated with AD risk or onset in these other two tissues. Given low overlap (Fig. 1) in individuals who have proteomic data across tissues, this was used as an internal validation. By leveraging across-tissue data, any tissue-specific signal will not replicate. One caveat of using the multi-tissue data is that not all proteins passed QC across all three tissues. Among the 12 proteins associated with AD status in brain, only 6 were found in both CSF and plasma. Of these, 5 proteins (SMOC1, HGF, FSTL1, UBC9, and NET1) were associated with AD status or age at onset

in both CSF and plasma data ($P < 0.05$, Supplementary Table 2), which represents an enrichment of 333 fold ($P = 5.8\times10^{-13}$) to what would be expected by chance.

To externally replicate these findings, we used the merged mass-spectrometry brain data (MassSpec Joint) that includes 10,078 proteins from 415 AD patients and 194 controls, and performed association analyses with AD status. As the proteomic data available in these studies were generated using a different platform, we were not able to test all 12 proteins that were significant in our discovery data. Of the nine proteins that were present in these datasets, 8 replicated (Midkine, SMOC1, CgA, HGF, NRX1B, UBC9, NET1, and SAP) with $P < 0.05$ and in the same direction of effect. This represents an enrichment of 35 fold to what would be expected by chance ($P = 1.3\times10^{-12}$). In addition, to confirm that our results were not false positives due to the joint analysis that included all 6 studies, we performed additional analyses in each study (Johnson et al,[15] Higginbotham et al,[12] and Wingo et al.[16]). Individual study analyses also provided enrichments of 25-34 fold (Supplementary Table 1). We also found a significant correlation in the effect size for the association of the proteins with AD status between our discovery results and the merged replication results (MassSpec Joint) ($P < 3.6\times10^{-3}$; Supplementary Fig. 3a). Together, these results indicate that our identified brain proteomic signature replicates in external independent samples and is extremely robust across orthogonal proteomics platforms.

CSF proteomic profiles for sporadic AD

In CSF, 117 proteins were associated with clinical AD status after Bonferroni correction (Fig. 2a, Supplementary Table 3). Of these 117 proteins, 78 passed QC in brain and plasma tissues, and 27 proteins (including ERK-1 and LRRK2) replicated in both tissues (138-fold enrichment, $P = 3.3\times10^{-50}$). An additional 44 proteins replicated in brain and 16 in plasma. To externally replicate our identified proteins in CSF, we downloaded and analyzed Alzheimer's Disease Neuroimaging Initiative (ADNI) multiple reaction monitoring (MRM) proteomic data containing 320 proteins in 263 samples. In addition, we obtained results based on BioFinder OLINK data of 201 proteins in 576 samples presented by Whelan et al.,[11] and from the mass-spectrometry-based Emory-ADRC study that includes 2,875 proteins in just 40 samples presented by Higginbotham et al.[12] Of the 117 CSF proteins identified in our study, 90 were present in these external datasets. Of these, 39 proteins (including 14-3-3, Calcineurin, SMOC1, GFAP, SPP1, and Peroxiredoxin-1) replicated in the same direction (14- to 34-fold enrichments, $P \leq 4.4\times10^{-5}$). It is important to mention that the major overlap in the number of proteins with our data is the Emory-ADRC study, which only includes 40 samples. Therefore, the power to replicate the initial findings is limited. We expect that a larger number of proteins will replicate in larger studies.

Several studies have demonstrated that up to 30% of cognitively normal elderly individuals could be pre-symptomatic for AD[17] and that other neurodegenerative diseases can masquerade, clinically, as AD dementia.[18] Therefore, clinically defined case-control status may not be the best phenotype for novel biomarker discovery.[19] It has been proposed that biomarker-based categorization provides a more powerful approach to identify proteins altered in AD. CSF Aβ42 and p-tau levels are one of the best fluid biomarkers identified to date for distinguishing pathology-free controls from AD dementia and several studies have demonstrated that CSF p-tau/Aβ42 ratio is a marker not only for AD status but also for predicting AD progression from normal to dementia within 5 years.[3] As we had access to CSF p-tau/Aβ42 for most samples with CSF (689 out of 720), we also performed a regression analysis of protein levels considering p-tau/Aβ42 ratio as a predictor. We found 92 proteins that were significant for p-tau/Aβ42 ratio at Bonferroni-corrected threshold. Of the 117 proteins associated with clinical AD status, 74 were significant for CSF p-tau/Aβ42 at Bonferroni-corrected threshold and the remaining were nominally significant. In fact, we found a very strong correlation (R^2=0.86 and $P < 1.0\times10^{-16}$; Supplementary Fig. 4) of the effect across all 713 QCed proteins between the two analyses. This indicates that using case-control status for the Knight ADRC is highly accurate and leads to the similar results as using biomarker-defined case-control status

Plasma proteomic profiles for sporadic AD

In plasma, 26 proteins were associated with sporadic AD status after Bonferroni correction (Fig. 2a, Supplementary Table 4). Similar to previous analyses, we leveraged the multi-tissue data to replicate these findings. Of the 26 plasma proteins associated with AD status, 16 passed QC in brain and CSF and seven proteins (including ERK-1, CDON, and SHC1) replicated (175-fold enrichment, $P= 6.8\times10^{-15}$). To externally replicate our findings, we downloaded the AddNeuroMed SOMAscan 1.1K proteomic data that was processed and deposited by Sattlenecker et al[13] and performed differential analysis in 320 individuals with AD and 194 controls. Out of 26 proteins, we were able to test 19 in this dataset and 9 proteins (including CAMK2D and HMG-1) replicated (18.9-fold enrichment, $p = 2.8\times10^{-10}$).

In summary, we have identified 8, 39, and 9 proteins that are associated with AD status and replicated in several independent cohorts using orthogonal technologies in brain, CSF and plasma, respectively. These proteins likely represent only a subset of proteins that could be associated with AD status, as not all proteins identified in our study were assayed in the replication datasets and most of the replication datasets had smaller sample sizes than our discovery data, providing limited power. We also leveraged multi-tissue data to replicate the single-tissue findings. Sometimes, it may not be possible to use external datasets for replication, therefore we performed an enrichment test to determine whether the proteins that showed an internal cross-tissue replication would also replicate in other studies. Our analyses indicate that proteins identified in each tissue and supported by the two remaining tissues were more likely to

replicate in external independent datasets (15- to 40-fold enrichments, $P \leq 3.63\times10^{-3}$, Supplementary Table 5), suggesting that multi-tissue proteomic data may be used as a viable replication strategy.

TREM2 risk variant carriers

Our group and other, identified several rare coding variants in *TREM2* that increase risk of AD by almost two fold, making *TREM2* the second strongest genetic risk factor for sporadic AD after *APOE*.[1,20-23] Multiple *TREM2* risk variants have been identified, but it has been proposed that all *TREM2* AD-risk variants cause a partial loss of function[24]. Given the low frequency of these variants, performing separate analysis for each specific variant would not provide enough statistical power. For these reasons, we combined all *TREM2* variant carriers in these analyses. We generated proteomic data from 21, 47, and 131 *TREM2* variant carriers in brain, CSF, and plasma, respectively (Table 1). To identify multi-tissue proteomic signatures of individuals carrying AD-risk variants in *TREM2*, we compared the protein levels of *TREM2* variant carriers with both cognitively normal individuals and individuals who were diagnosed with AD dementia, but did not carry any TREM2 or autosomal dominant variant. This is the first time a proteomic profile for *TREM2* variant carriers has been generated.

In the brain, 9 proteins (including α-Synuclein) showed differential abundance levels in *TREM2* variant carriers compared to cognitively normal individuals at Bonferroni-corrected threshold (Fig. 3a; Supplementary Table 6). In addition, 23 proteins (including LRRK2) were associated with AD status after multi-test correction for *TREM2* risk variant carriers vs. AD (Supplementary Table 7). From the genetic data available for the replication datasets, we found 4 *TREM2* variant carriers in Mayo, 7 in MSBB, and 8 in ROSMAP. This low number did not provide any statistical power to support a replication analysis. As we demonstrated, our multi-tissue study design is a viable alternative approach to identify proteins that would replicate in external datasets, and we leveraged our data to identify those proteins that replicate across tissues. Out of these 27 unique *TREM2*-associated proteins (combining 9 and 23 proteins), 11 passed QC in both CSF and plasma, and 5 (ALT, α -Synuclein, MIS, LRRK2, and PAFAH beta subunit) replicated in both tissues. This represents a 74-fold enrichment ($p=7.53\times10^{-9}$) to what would be expected by chance.

In CSF, our analyses identified a total of 38 unique proteins, among which 31 were associated with *TREM2* risk variant carriers vs. cognitively normal individuals and 10 for the *TREM2* vs. AD, after multiple test correction (Supplementary Tables 8-9). Out of these 38 proteins, 20 passed QC in the other tissues, and 7 (14-3-3E, 14-3-3 protein zeta/delta, Somatostatin-28, SMOC1, Ubiquitin+1, QORL1 and calcineurin) replicated across tissues (Supplementary Tables 8-9). This represents a 73-fold enrichment ($p=7.19\times10^{-12}$) to what would be expected by chance.

In the plasma proteomic data, we identified a total of 69 proteins, among which 65 and 7 showed differential abundance levels in *TREM2* variant carriers compared to cognitively normal individuals and to individuals who were diagnosed with AD dementia, respectively (Supplementary Tables 10-11). Among the 41 proteins that passed QC in the brain and CSF, 21 proteins (including bone proteoglycan II, PAPP-A, ERK-1, suPAR and VCAM-1) replicated, which represents a 122-fold enrichment ($p=5.47\times10^{-38}$) to what would be expected by chance.

Autosomal dominant AD status

Although most AD cases are considered sporadic and manifest after the age of 65,[24] around 1-3% of AD cases show an autosomal dominant (ADAD) inheritance pattern, often with onset before age 65.[25] Pathogenic variants in *APP*, *PSEN1* and *PSEN2* have been identified as the cause of ADAD.[9] We generated proteomic data from the parietal cortex of 24 ADAD gene variant carriers (19 individuals with *PSEN1*, 1 with *PSEN2*, and 4 with *APP* variants) recruited from the DIAN and the Knight ADRC studies. We identified 109 proteins with differential abundance in ADAD mutation carriers compared to cognitively normal individuals with no significant brain pathology, at Bonferroni corrected threshold (Supplementary Fig. 5). In order to validate these findings, we analyzed whether these 109 proteins were also associated with ADAD status in CSF from 289 carriers and 184 non-carriers from the DIAN study. Due to the limited amount of CSF samples for these subjects, we were unable to perform proteomic discovery in sporadic AD or *TREM2* variant carriers. From those 109 proteins identified in brain, 106 passed QC in CSF proteomic data and 17 were associated with ADAD in CSF and in the same direction (Fig. 4, Supplementary Table 12), which represents a 6.4-fold enrichment ($p=1.36\times10^{-9}$) to what would be expected by chance.

As presented earlier, we identified 12 proteins associated with sporadic AD status in brain tissue (Supplementary Table 2). We also sought to determine if the proteins associated with sporadic AD status showed similar differential abundance in ADAD mutation carriers. We found that most of the proteins associated with sporadic AD brains displayed even stronger effect size when comparing ADAD mutation carriers to controls (Supplementary Table 13). The proteins associated with sporadic AD status showed 39% higher effect sizes in ADAD brain samples on average ($P = 3.8\times10^{-5}$; Fig. 4). For example, SMOC1 showed a significant association AD vs. control (Effect = 0.04: $P=3.1\times10^{-6}$) but also for ADAD vs. CO (Effect = 0.13; $P=2.3\times10^{-6}$). As presented earlier, SMOC1 has also been found to be associated in sporadic AD status in both CSF ($P=8.4\times10^{-29}$) and plasma (P=0.002), suggesting that it could be used to create a new prediction model for AD, independent of Aβ and tau.

Tissue-specific Prediction Models

Our analyses identified tissue-specific proteomic signatures for sporadic AD and *TREM2* risk-variant carriers. Here, we used the proteins that replicated in external datasets (for AD status) or across tissues (for *TREM2* variant carriers and ADAD) to create prediction models. To assess the specificity and selectivity of our prediction models we computed receiver operator characteristic (ROC) curve and area under the curve (AUC) using the R package pROC. Age at measurement and sex were included as covariates. We also performed analysis by adding *APOE e4* status as a covariate. In sporadic AD cases, these prediction models were examined for both the discovery and replication datasets.

In brain tissue, our prediction model based on the 8 proteins that replicated in our analysis (Supplementary Tables 1, 14) led to an AUC of 0.84 in the discovery and an AUC of 0.99 in the replication cohort (Fig. 2b). In CSF, we found 39 proteins associated with AD status that replicated in external datasets (Supplementary Table 3). A prediction model including these proteins led to an AUC of 0.90 in the replication and of 0.89 in the discovery cohort (Fig. 2b). As the number of proteins is too large to generate a prediction model that could be translated to the clinic, we performed the stepwise model selection to identify the minimum set of proteins that capture the same information as the 39 identified in our study. We found a panel of 12 proteins that provided accuracy in distinguishing clinically defined AD patients from controls almost as high as all 39 proteins and led to an AUC of 0.88 in the discovery and 0.999 in replication data. We compared our prediction model to CSF p-tau/Aβ42, known and validated biomarkers. In our dataset the CSF p-tau/Aβ42 ratio led to an AUC of 0.81, which is significantly lower than our prediction model ($P = 2.4\times10^{-6}$). Using the same approach for plasma, the 9 proteins identified and replicated in an external dataset (Supplementary Table 4, 14) led to an AUC of 0.79 in both discovery and replication datasets, which was not statistically different from the AUC with CSF p-tau/Aβ42 ratio (AUC=0.82; P>0.05). The prediction model based on each externally replicated protein is similar between the discovery and replication data (Supplementary Fig. 6).

We also created prediction models that could distinguish *TREM2* variant carriers from non-carriers in both sporadic AD cases and controls. Therefore, we included the proteins that were differentially abundant between *TREM2* risk variant carriers when compared not only to AD cases but also to controls. Due to a lack of external datasets, we included only those proteins that replicated across tissues, as explained above. In CSF, the prediction model that included 7 proteins (Supplementary Tables 8-9) resulted in an AUC of 0.79 when comparing *TREM2* risk variant carriers to controls. The same proteins showed an AUC of 0.84 for *TREM2* risk variant carriers compared to AD cases (Fig. 3b). CSF p-tau/Aβ42 levels have been shown to be a very good biomarker to distinguish AD cases vs controls, but no previous studies examined how CSF p-tau/Aβ42 ratio provides prediction for *TREM2* variant carriers. In this study, CSF p-tau/Aβ42

showed an AUC of 0.74 for *TREM2* variant carriers vs AD cases and AUC of 0.53 for TREM2 risk variant carriers vs cognitively normal individuals. Both AUC values are significantly lower than those from our *TREM2*-associated prediction model with 7 proteins ($P < 1.6\times10^{-5}$; Fig. 3b).

In plasma, the 21 proteins included in the model (Supplementary Tables 10-11) led to an AUC of 0.93 in differentiating *TREM2* risk variant carriers from controls, while the CSF p-tau/Aβ42 ratio led to a significantly lower AUC of 0.69 ($P = 1.1\times10^{-3}$). Similarly, in differentiating *TREM2* risk carriers from other AD cases, the same 21 proteins led to an AUC of 0.90, which is significantly higher ($P = 1.5\times10^{-4}$) than the AUC with the CSF p-tau/Aβ42 ratio (AUC=0.63). As the number of proteins is large, we performed a stepwise model selection and found a subset of 9 proteins that provided AUCs of 0.89 and 0.88 to discriminate *TREM2* variant carriers from cognitively normal individuals and from individuals with AD dementia, respectively (Fig. 3b). The prediction models including age, sex and *APOE e4* status as covariates provided similar performance (Supplementary Fig. 7).

We also leveraged the 17 proteins that were found to be associated with ADAD status and in the same direction in brain and CSF (Supplementary Table 12) to create potential prediction models for distinguishing ADAD mutation carriers from non-carriers. In brain data, the model with these 17 proteins provided an AUC of 1, which is significantly higher than the model based on age alone (AUC = 0.76; $P = 9.9\times10^{-3}$). In CSF data, the same 17 proteins provided a higher AUC value than the model with age alone (AUC = 0.87 vs 0.53, $P < 2.2\times10^{-16}$; Fig. 4).

Pathway Enrichment

Finally, we wanted to determine if the proteins identified in our analyses were enriched in common functional pathways. Functional enrichment analysis was performed with Enrichr.[26] As expected, the AD pathway was significant in CSF in both the sporadic AD (FDR = 1.9×10^{-3}) and *TREM2* variant-specific analyses (FDR = 5.8×10^{-3}, Supplementary Table 15). The proteins that are part of this pathway that were identified in our analyses include APOE, calcineurin (*PPP3R1* and *PPP3CA*), and MAPK3 (Fig. 5, Supplementary Fig. 8). APOE is the strongest and most common genetic risk factor for AD,[27] and individuals with the *APOE e4* allele have lower CSF Aβ42 levels[27] and lower Aβ42 clearance.[28,29] Genetic variants in calcineurin have been associated with higher CSF p-tau levels and earlier age at onset.[30] *MAPK3* has also been reported to be involved in AD pathology,[31-33] likely by affecting tau phosphorylation. In any biomarker discovery study, it is often difficult to determine whether the proteins

identified are part of a causal pathway or just a product of the disease. Several facts strongly suggest that many of the proteins identified in this study are, in fact, causal. As mentioned, *APOE* is known to be part of the causal AD pathway, and calcineurin and *MAPK3* have recently been reported as part of the causal AD pathway by pQTL and Mendelian randomization analyses[5] .

Several proteins that are part of the Parkinson disease pathway, including α-synuclein, LRRK2, granulin, and UCHL1, were also found to be dysregulated in CSF and plasma for the sporadic AD and *TREM2* analyses (FDR < 3.4×10^{-3}, Supplementary Table 15). On autopsy, around 30% of the AD cases, including autosomal dominant AD, present with Lewy bodies, which are deposits of α-synuclein.[34] Those reports, together with our analyses, indicate that PD pathology shares similarities with AD pathology. Similar to α-synuclein, LRRK2 also showed a strong association with autosomal dominant AD (P = 7.7×10^{-4}) and *TREM2* (P = 9.3×10^{-6}). The *GRN* gene, which encodes the granulin protein, was initially associated with frontotemporal dementia,[35,36] but recent, large GWAS have also found *GRN* in both AD[37] and PD.[38]

Granulin, implicated in wound healing[39] as a part of the innate immune response pathway, was also found to be enriched in the proteomic analyses for sporadic AD in CSF (FDR = 6.9×10^{-9}) and plasma (FDR = 2.1×10^{-3}), as well as the CSF TREM2-specific analyses (FDR = 1.1×10^{-3}). Other dysregulated proteins identified in our analyses that are also part of this pathway include SHC1, MAPK3, ITGB1, and SPP1, among others. SPP1 has recently been implicated in microglia activation and the AD pathway.[40] Similar to *SPP1*, *ITBG1* is a microglia gene and has been shown to be differentially expressed in the hippocampus and peripheral blood mononuclear cells (PBMC) of AD cases,[41] important in microglia activation,[42] and part of the causal pathway in network analyses.[43] Recent studies have also demonstrated that meningeal lymphatics affect microglia and AD risk.[44] Our analyses also found several endothelial-specific proteins (ERK-1, SHC1, and BCAM).

The 17 proteins that were associated with ADAD status in both brain and CSF in the same direction, were also enriched for proteins part of the Alzheimer disease pathway (p<1×10-4) and

the cellular response to chemical stimulus pathway (go:0.0070887; p=0.034), which includes, among others, MIF, a pro-inflammatory cytokine involved on involved in the innate immune response; LILRB; and CD22 also part of the immune response pathway. IDE is involved in the cellular breakdown of insulin and has been reported to be involved in the degradation and clearance of naturally secreted amyloid beta-protein by neurons and microglia.

In summary, the proteins dysregulated in our analyses are not randomly distributed across functional groups; they are enriched in specific pathways known to be implicated in AD and other pathways (PD, immune response) that may be instrumental to AD pathophysiology and may represent new therapeutic targets. Indeed, our analyses indicate that the proteins identified here are not only dysregulated in AD but also play a causal role.

Discussion

This is the first large-scale, multi-tissue proteomic characterization of sporadic and genetically defined AD cases (*TREM2* and Mendelian cases). We created a web portal (http://ngi.pub:3838/ONTIME_Proteomics/) to facilitate the exploration of our analyses and further investigation into individual protein abundance levels across disease status or sex (Supplementary Fig. 9). In this study, we obtained proteomic measures from Knight ADRC and DIAN cohorts and identified proteomic profiles for sporadic AD, *TREM2* variant carriers, and autosomal dominant AD cases in three tissues. These proteomic profiles replicated in independent datasets and across tissues, which were used to create tissue-specific prediction models and to identify novel causal proteins and pathways for sporadic and genetically defined AD cases. We created and validated tissue-specific prediction models using proteins identified in CSF and plasma that were as good as, or better than, the current gold standard antibody-based biomarkers for AD risk. Having new prediction models in CSF and plasma that are independent from Aβ and tau might be relevant for clinical trials and therapies that target those molecules, as biomarkers that do not rely on the targe protein may be needed. We also demonstrated that there are common proteins associated with AD status across tissues, which has important implications for the identification and validation of AD biomarkers in future studies.

This study also identified new proteins and pathways implicated in sporadic AD and individuals with specific genetic profiles. While validation of some of our findings will require additional follow-up studies, these results highlight the need for multi-tissue proteomics to fully understand the biology of AD and create tissue-specific prediction models for individuals with specific genetic profiles, ultimately supporting its utility in generating clinically useful biomarker arrays. This study indicates that once individuals with specific genetic profiles are identified, it is possible to create customized prediction models and identify proteins implicated in disease, an instrumental step toward creating individualized, specific disease risk evaluation and treatment.

Methods And Materials

Study Participants

This study included the brain (N=360), CSF (N=717), and plasma (N=490) data from the Knight ADRC[3] and the Dominantly Inherited Alzheimer Network (DIAN)[9] cohorts. The recruited individuals were

evaluated by Clinical Core personnel of the Knight ADRC.

For brain samples, brain autopsy was performed by the Knight ADRC Neuropathology Core and AD status was determined by postmortem neuropathological analysis. Brain tissues were collected from fresh frozen human parietal lobes. Neuropathological phenotypes, including Braak tau, CERAD Aβ, α-synuclein pathology, postmortem interval (PMI), age at onset, age at death and brain weight, were obtained for all brain samples. The brain data included 24 individuals carrying autosomal dominant AD (ADAD) mutations, out of which 18 were from the DIAN cohort. Among these ADAD individuals, 19, 1, and 4 carried pathogenic mutations in *PSEN1*, *PSEN2*, and *APP*, respectively.

Among individuals with CSF and plasma data, AD cases corresponded to those with a diagnosis of dementia of the Alzheimer's type (DAT) using criteria equivalent to the National Institute of Neurological and Communication Disorders and Stroke-Alzheimer's Disease and Related Disorders Association for probable AD,[6,456,43] and AD severity was determined using the Clinical Dementia Rating (CDR®)[46] at the time of lumbar puncture (for CSF samples) or blood draw (for plasma samples). Controls received the same assessment as the cases but were non-demented (CDR=0). CSF and blood for plasma were collected in the morning after an overnight fast, aliquoted and stored at -80°C until assayed.[3,9] CSF Aβ and tau levels were measured as explained previously.[3] The Institutional Review Board of Washington University School of Medicine in St. Louis approved the study and research was performed in accordance with the approved protocols.

Proteomic Data

For deep omics characterization in brain, CSF, and plasma tissues, we quantified the level of 1,305 proteins using a multiplexed, single-stranded DNA aptamer assay developed by SomaLogic.[47] The assay covers a dynamic range of 10^8 and measures all three major categories: secreted, membrane, and intracellular proteins. The proteins cover a wide range of molecular functions and include proteins known to be relevant to human disease. Aliquots of gray matter homogenate (150 µl) of tissue were provided to the Genome Technology Access Center at Washington University in St. Louis for protein measurement. As previously described by Gold et al,[17] modified single-stranded DNA aptamers are used to bind specific protein targets, which are then quantified by a DNA microarray. Protein concentrations are quantified as relative fluorescent units (RFU) of intensity in this DNA microarray.

Quality control (QC) was performed at the sample and aptamer levels using control aptamers (positive and negative controls) and calibrator samples. At the sample level, hybridization controls on each plate

were used to correct for systematic variability in hybridization. The median signal over all aptamers was used to correct for within-run technical variability. This median signal was assigned to different dilution sets within each tissue. For brain and CSF samples, a 20% dilution rate was used. For plasma samples, three different dilution sets (40%, 1%, and 0.005%) were used.

As described in detail,[5] we performed additional QC by identifying and removing protein/analyte outliers by applying the following four criteria using R. 1) Minimum detection filtering. The limit of detection (LOD) was computed based on negative controls. If the average expression of an analyte in a sample was found to be less than its LOD in more than 15% of total sample size, this sample was marked as an outlier and excluded. 2) Scale factor difference. Scale factor difference was calculated as the maximum value of the absolute difference between the median expression of analytes per plate and calibration scale factor. If the maximum difference was greater than 0.5, the analyte was excluded. 3) Coefficient of variation (CV). The CV for each aptamer was calculated as the standard deviation divided by the mean of the protein levels in calibrators. If the median coefficient of variation for a particular analyte was greater than 0.15, this analyte was excluded. 4) Interquartile range (IQR). If more than 15% of the log10 transformed analyte values are located outside of either end of a 1.5-fold of IQR, this analyte was marked as an outlier and excluded. In addition, if more than 15% of the transformed analyte values in a particular sample are located outside a 1.5-fold of IQR, this sample was marked as an outlier and excluded. Analytes and samples that remained after applying these 4 criteria were used for the downstream statistical analysis.

Differential Abundance Analysis

To obtain proteomic signatures of sporadic AD status, *TREM2* risk variant carriers, and autosomal dominant AD (ADAD) status, we performed differential abundance analysis by using log10-transformed protein levels as an outcome in a linear regression model. In all three tissues, sporadic AD status and *TREM2* variant carrier status were considered as a main predictor. In brain tissue, ADAD status was also considered. In each tissue, we performed surrogate variable analysis (SVA) while including status and age as covariates in a null hypothesis model to remove batch effects in our proteomics data (17 batches in brain, 50 batches in CSF and 27 batches in plasma data) and correct for other unmeasured heterogeneity.[14] The number of resulting surrogate variables were 10, 32, and 14 in brain, CSF, and plasma, respectively. Age at death or at measurement (in all regression models except for ADAD-specific analysis), sex and the resulting surrogate variables were included as covariates. In ADAD analysis in brain tissue, sex was excluded from covariates as control group was older than ADAD individuals.

In addition, we performed analyses using age-at-onset (AAO) and AD neuropathology characteristics (Braak neurofibrillary tangle scores and CDR at death) for brain data, AAO and CSF pTau/Aβ42 ratio for

CSF data, and AAO for plasma data, while including the same covariates. For AAO, we performed survival analysis while considering age, sex and surrogate variables as covariates (Supplementary Fig. 1). We created a survival object using R function Surv and performed a Cox proportional hazards regression model using the coxph function. In addition, we examined the consistency between effect sizes of AD status and AD neuropathology measures through the scatter plots. We performed correlation tests using cor.test in R to test association between effect sizes with Pearson's product moment correlation coefficient and two-sided alternative hypothesis. In addition, we performed Fisher's exact test for the same direction.

We obtained the minimum number of principal components (PCs) that cumulatively explain 95% of the variance for each tissue after QC. The number of PCs is 75, 169, and 230 in brain, CSF, and plasma data, respectively. We considered a Bonferroni-corrected threshold as 0.05 divided by this number of PCs. The thresholds corresponded to 0.67×10^{-4} in brain, 2.96×10^{-4} in CSF, and 2.21×10^{-4} in plasma. When we applied Bonferroni correction and false discovery rate (FDR), we found that the use of these Bonferroni-corrected thresholds usually provided fewer significant results and is therefore more conservative than the use of FDR (Supplementary Table 16). Because of this, we chose to apply Bonferroni correction.

Replication Strategies

To internally validate our identified proteins in each tissue, we examined which proteins would be associated in the remaining tissues at the nominal significance threshold ($P < 0.05$). In addition, to externally replicate our sporadic AD findings within the same tissue, we downloaded multiple publicly available proteomic datasets.

For brain tissue, we downloaded the mass-spectrometry data that were processed and deposited by Johnson et al[15] for following 6 studies: the Adult Changes In Thought (ACT), Banner Sun Health Research Institute (BANNER), Baltimore longitudinal study of aging (BLSA), Mayo Clinic (MAYO), Mount Sinai Brain Bank (MSBB), the Religious Orders Study and the Memory and Aging Project (ROSMAP). We combined these brain proteomics data from all 6 studies (resulting in a total of 10,078 proteins measured in 415 AD patients and 194 controls) and performed SVA to account for batch effects and unmeasured heterogeneity. Then we performed differential abundance levels of AD status jointly while considering age, sex and 11 surrogate variables as covariates. In addition, to confirm that our results were not false positives due to the joint analysis merging of all 6 studies, we used the results presented by Johnson et al[15] that used ACT, BANNER, BLSA, MSBB, results by Higginbotham et al[12] that used DLPFC, and results by Wingo et al[16] that used ROSMAP.

For CSF tissue, we obtained multiple reaction monitoring ADNI data and performed differential analysis for 320 proteins in 188 AD patients and 75 controls while considering age, sex and 7 surrogate variables as covariates. In addition, we used the results based on Emory-ADRC mass-spectrometry data from Higginbotham et al[12] and the results based on BioFinder OLINK data from Whelan et al.[11] For plasma tissue, we downloaded the cleaned SOMAscan 1.1K proteomic data from the AddNeuroMed study.[13] After excluding 166 individuals with mild cognitive impairment, we performed differential abundance analysis of 320 AD patients and 194 controls, while including sex, age, batch effects, and APOE status as covariates. For ADAD status, only high throughput proteomic screening was performed on brain tissue. To replicate these proteins that were associated with ADAD status in brain, we performed analysis in CSF from 289 ADAD mutation carriers and 184 non carriers from the DIAN study[9], while including sex, age, and batch effects as covariates.

To compute the fold-enrichment of replication to what would be expected by chance, we used the Binomial distribution. Under the null hypothesis of no enrichment, the expected number of replicated proteins is the number of available proteins for testing times the significance threshold (0.05 x 0.05 in across-tissue replication and 0.05 x 0.5 in external replication). We computed the enrichment as the ratio of the observed replications by the expected replications and the p-value based on the Binomial probability.

ADNI

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD).

Prediction Models

To obtain tissue-specific prediction models, we performed logistic regression models considering multiple proteins as main predictors and sporadic AD and *TREM2* variant carrier status as an outcome, while including sex, age, and with and without *APOE e4* allele status as covariates. In sporadic AD status, externally validated proteins were used for both discovery and replication datasets. In brain, we used the combined mass-spectrometry brain proteomics data from ACT, BANNER, BLSA, MSBB, MAYO, and

ROSMAP for replication. In CSF, we downloaded the Emory-ADRC mass-spectrometry data that were processed and deposited by Higginbotham et al.[12] and used for replication. In plasma, we used cleaned data from the AddNeuroMed[13] SOMAscan 1.1K proteomic dataset for replication. In *TREM2* variant carrier status modeling, we used proteins validated across tissues for discovery data.

When there were more than 10 proteins, we also performed stepwise regression analysis to reduce the number of proteins by selecting the best model by Akaike information criterion (AIC) using step function in R. We considered both forward and backward selection and chose the model with fewer proteins when there were two competing models. The well accepted CSF biomarker using p-Tau/Aβ42 ratio was considered as a gold standard for comparison. Receiver operator characteristic (ROC) curves and areas under the curves (AUC) were computed using the R package pROC V1.12.1. The roc.test function within the same package was used to compare AUC values for two models (such as AUC based on the identified proteins vs. AUC based on p-Tau/Aβ42 ratio).

Pathway Enrichments

Functional enrichment analysis was performed with Enrichr.[26] For sporadic AD findings, the genes that target our proteins identified and validated internally or externally (9, 42 and 14 genes in brain, CSF, and plasma, respectively) were used as an input for enrichment analysis. For *TREM2* carrier status, we considered 6, 10 and 21 genes that replicated across tissues, in brain, CSF, and plasma, respectively. Among multiple gene-set libraries, KEGG, Reactome, Panther pathways and GO biological process were considered. The significance of functional enrichment was reported as the p-value of Fisher's exact test, followed by Benjamini–Hochberg adjustment for false discovery rates (FDR) in testing multiple hypotheses. We considered results with FDR < 0.05 as significant and included them for creating the dot chart and tile plots to graphically display our findings.

Tables

Table 1: Summary characteristics of participants with proteomic measures in the Knight ADRC and DIAN cohorts

Tissue	Status	Sample size (N)	% Female	Age (mean ± SD)
Brain	CO	25	61.72	88.24 ± 8.85
	AD	290	33.33	83.98 ± 8.83
	ADAD	24	76.00	55.67 ± 14.58
	TREM2	21	57.14	82.57 ± 7.62
CSF	CO	494	55.26	73.15 ± 6.43
	AD	176	46.02	74.60 ± 7.02
	TREM2	47	44.68	74.00 ± 6.48
Plasma	CO	254	57.48	71.53 ± 7.31
	AD	105	37.14	72.59 ± 7.67
	TREM2	131	64.89	74.98 ± 8.17

CO = healthy control; AD = sporadic AD cases; ADAD = Autosomal dominant AD; *TREM2* = AD-risk variant (p.E151K, p.H157Y, p.L211P, p.R136Q, p.R163Q, p.R47H, p.R62H, p.T96K) carriers in *TREM2*; CSF = cerebrospinal fluid.

References

1. Guerreiro, R. *et al.* TREM2 variants in Alzheimer's disease. *N Engl J Med* **368**, 117-27 (2013).

2. Karch, C.M., Cruchaga, C. & Goate, A.M. Alzheimer's disease genetics: from the bench to the clinic. *Neuron* **83**, 11-26 (2014).

3. Fagan, A.M. *et al.* Cerebrospinal fluid tau/beta-amyloid(42) ratio as a prediction of cognitive decline in nondemented older adults. *Arch Neurol* **64**, 343-9 (2007).

4. Mueller, S.G. *et al.* Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimers Dement* **1**, 55-66 (2005).

5. Yang, C. *et al.* Genomic atlas of the proteome from brain, CSF and plasma prioritizes proteins implicated in neurological disorders. *Nat Neurosci* (2021).

6. McKhann, G. *et al.* Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34**, 939-44 (1984).

7. Association, A.s. 2020 Alzheimer's disease facts and figures. *Alzheimers Dementia* (2020).

8. Jin, S.C. *et al.* Coding variants in TREM2 increase risk for Alzheimer's disease. *Hum Mol Genet* **23**, 5838-46 (2014).

9. Bateman, R.J. *et al.* Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *N Engl J Med* **367**, 795-804 (2012).

10. Bateman, R.J. *et al.* Autosomal-dominant Alzheimer's disease: a review and proposal for the prevention of Alzheimer's disease. *Alzheimers Res Ther* **3**, 1 (2011).

11. Whelan, C.D. *et al.* Multiplex proteomics identifies novel CSF and plasma biomarkers of early Alzheimer's disease. *Acta Neuropathol Commun* **7**, 169 (2019).

12. Higginbotham, L. *et al.* Integrated proteomics reveals brain-based cerebrospinal fluid biomarkers in asymptomatic and symptomatic Alzheimer's disease. *Sci Adv* **6**(2020).

13. Sattlecker, M. *et al.* Alzheimer's disease biomarker discovery using SOMAscan multiplexed protein technology. *Alzheimers Dement* **10**, 724-34 (2014).

14. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E. & Storey, J.D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882-3 (2012).

15. Johnson, E.C.B. *et al.* Large-scale proteomic analysis of Alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. *Nat Med* **26**, 769-780 (2020).

16. Wingo, A.P. *et al.* Shared proteomic effects of cerebral atherosclerosis and Alzheimer's disease on the human brain. *Nat Neurosci* **23**, 696-700 (2020).

17. Vlassenko, A.G. *et al.* Imaging and cerebrospinal fluid biomarkers in early preclinical alzheimer disease. *Ann Neurol* **80**, 379-87 (2016).

18. Guo, T. *et al.* Association of CSF Abeta, amyloid PET, and cognition in cognitively unimpaired elderly adults. *Neurology* **95**, e2075-e2085 (2020).

19. Hampel, H. *et al.* Developing the ATX(N) classification for use across the Alzheimer disease continuum. *Nat Rev Neurol* (2021).

20. Benitez, B.A. *et al.* TREM2 is associated with the risk of Alzheimer's disease in Spanish population. *Neurobiol Aging* **34**, 1711 e15-7 (2013).

21. Neumann, H. & Daly, M.J. Variant TREM2 as risk factor for Alzheimer's disease. *N Engl J Med* **368**, 182-4 (2013).

22. Sims, R. *et al.* Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nat Genet* **49**, 1373-1384 (2017).

23. Deming, Y. *et al.* The MS4A gene cluster is a key modulator of soluble TREM2 and Alzheimer's disease risk. *Sci Transl Med* **11**(2019).

24. Ulrich, J.D. & Holtzman, D.M. TREM2 Function in Alzheimer's Disease and Neurodegeneration. *ACS Chem Neurosci* **7**, 420-7 (2016).

25. Hsu, S. *et al.* Discovery and validation of autosomal dominant Alzheimer's disease mutations. *Alzheimers Res Ther* **10**, 67 (2018).

26. Chen, E.Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).

27. Cruchaga, C. *et al.* Cerebrospinal fluid APOE levels: an endophenotype for genetic studies for Alzheimer's disease. *Hum Mol Genet* **21**, 4558-71 (2012).

28. Castellano, J.M. *et al.* Human apoE isoforms differentially regulate brain amyloid-beta peptide clearance. *Sci Transl Med* **3**, 89ra57 (2011).

29. Verghese, P.B. *et al.* ApoE influences amyloid-beta (Abeta) clearance despite minimal apoE/Abeta association in physiological conditions. *Proc Natl Acad Sci U S A* **110**, E1807-16 (2013).

30. Cruchaga, C. *et al.* SNPs associated with cerebrospinal fluid phospho-tau levels influence rate of decline in Alzheimer's disease. *PLoS Genet* **6**(2010).

31. Zhao, W.Q. *et al.* MAP kinase signaling cascade dysfunction specific to Alzheimer's disease in fibroblasts. *Neurobiol Dis* **11**, 166-83 (2002).

32. Swatton, J.E. *et al.* Increased MAP kinase activity in Alzheimer's and Down syndrome but not in schizophrenia human brain. *Eur J Neurosci* **19**, 2711-9 (2004).

33. Xiao, Y. *et al.* Tetrahydrocurcumin ameliorates Alzheimer's pathological phenotypes by inhibition of microglial cell cycle arrest and apoptosis via Ras/ERK signaling. *Biomed Pharmacother* **139**, 111651 (2021).

34. Ringman, J.M. *et al.* Neuropathology of Autosomal Dominant Alzheimer Disease in the National Alzheimer Coordinating Center Database. *J Neuropathol Exp Neurol* **75**, 284-90 (2016).

35. Van Deerlin, V.M. *et al.* Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. *Nat Genet* **42**, 234-9 (2010).

36. Cruchaga, C. *et al.* Association of TMEM106B gene polymorphism with age at onset in granulin mutation carriers and plasma granulin protein levels. *Arch Neurol* **68**, 581-6 (2011).

37. Bellenguez, C. *et al.* New insights on the genetic etiology of Alzheimer's and related dementia. *medRxiv*, 2020.10.01.20200659 (2020).

38. Chang, D. *et al.* A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet* **49**, 1511-1516 (2017).

39. Paushter, D.H., Du, H., Feng, T. & Hu, F. The lysosomal function of progranulin, a guardian against neurodegeneration. *Acta Neuropathol* **136**, 1-17 (2018).

40. Thrupp, N. *et al.* Single-Nucleus RNA-Seq Is Not Suitable for Detection of Microglial Activation Genes in Humans. *Cell Rep* **32**, 108189 (2020).

41. Ma, G. *et al.* Differential Expression of mRNAs in the Brain Tissues of Patients with Alzheimer's Disease Based on GEO Expression Profile and Its Clinical Significance. *Biomed Res Int* **2019**, 8179145 (2019).

42. Pan, X.D. *et al.* Microglial phagocytosis induced by fibrillar beta-amyloid is attenuated by oligomeric beta-amyloid: implications for Alzheimer's disease. *Mol Neurodegener* **6**, 45 (2011).

43. Li, Z., Xiong, Z., Manor, L.C., Cao, H. & Li, T. Integrative computational evaluation of genetic markers for Alzheimer's disease. *Saudi J Biol Sci* **25**, 996-1002 (2018).

44. Da Mesquita, S. *et al.* Meningeal lymphatics affect microglia responses and anti-Abeta immunotherapy. *Nature* **593**, 255-260 (2021).

45. Berg, L. *et al.* Clinicopathologic studies in cognitively healthy aging and Alzheimer's disease: relation of histologic markers to dementia severity, age, sex, and apolipoprotein E genotype. *Arch Neurol* **55**, 326-35 (1998).

46. Morris, J.C. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* **43**, 2412-4 (1993).

47. Gold, L. *et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* **5**, e15004 (2010).

Declarations

Acknowledgements

We thank all the participants and their families, as well as the many involved institutions and their staff.

Funding: This work was supported by grants from the National Institutes of Health (R01AG044546 (CC), P01AG003991(CC, JCM), RF1AG053303 (CC), RF1AG058501 (CC), U01AG058922 (CC), and R01AG057777 (OH)), and the Chuck Zuckerberg Initiative (CZI).

The recruitment and clinical characterization of research participants at Washington University were supported by NIH P30AG066444 (JCM), P01AG03991(JCM), and P01AG026276(JCM). O.H. is an Archer Foundation Research Scientist.

This work was supported by access to equipment made possible by the Hope Center for Neurological Disorders, the Neurogenomics and Informatics Center (NGI: https://neurogenomics.wustl.edu/)and the Departments of Neurology and Psychiatry at Washington University School of Medicine.

DIAN resources: Data collection and sharing for this project was supported by The Dominantly Inherited Alzheimer Network (DIAN, U19AG032438) funded by the National Institute on Aging (NIA),the Alzheimer's Association (SG-20-690363-DIAN), the German Center for Neurodegenerative Diseases (DZNE), Raul Carrea Institute for Neurological Research (FLENI), Partial support by the Research and Development Grants for Dementia from Japan Agency for Medical Research and Development, AMED, and the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), Spanish Institute of Health Carlos III (ISCIII), Canadian Institutes of Health Research (CIHR), Canadian Consortium of Neurodegeneration and Aging, Brain Canada Foundation, and Fonds de Recherche du Québec – Santé. This manuscript has been reviewed by DIAN Study investigators for scientific content and consistency of data interpretation with previous DIAN Study publications. We acknowledge the altruism of the participants and their families and contributions of the DIAN research and support staff at each of the participating sites for their contributions to this study

ADNI acknowledgement: Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

AMP-AD Acknowledgements: The results published here are in whole or in part based on data obtained from the AD Knowledge Portal (https://adknowledgeportal.org). Study data were provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago. Data collection was supported

through funding by NIA grants P30AG10161 (ROS), R01AG15819 (ROSMAP; genomics and RNAseq), R01AG17917 (MAP), R01AG30146, R01AG36042 (5hC methylation, ATACseq), RC2AG036547 (H3K9Ac), R01AG36836 (RNAseq), R01AG48015 (monocyte RNAseq) RF1AG57473 (single nucleus RNAseq), U01AG32984 (genomic and whole exome sequencing), U01AG46152 (ROSMAP AMP-AD, targeted proteomics), U01AG46161(TMT proteomics), U01AG61356 (whole genome sequencing, targeted proteomics, ROSMAP AMP-AD), the Illinois Department of Public Health (ROSMAP), and the Translational Genomics Research Institute (genomic). Additional phenotypic data can be requested at www.radc.rush.edu. Study data were provided through NIA grant 3R01AG046171-02S2 awarded to Rima Kaddurah-Daouk at Duke University, based on specimens provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, where data collection was supported through funding by NIA grants P30AG10161, R01AG15819, R01AG17917, R01AG30146, R01AG36836, U01AG32984, U01AG46152, the Illinois Department of Public Health, and the Translational Genomics Research Institute.

Conflict of interest

CC has received research support from: Biogen, EISAI, Alector and Parabon. The funders of the study had no role in the collection, analysis, or interpretation of data; in the writing of the report; or in the decision to submit the paper for publication. CC is a member of the advisory board of Vivid genetics, Halia Therapeutics and ADx Healthcare. There are no conflicts.

Dr. Fagan has received research funding from the National Institute on Aging of the National Institutes of Health, Biogen, Centene, Fujirebio and Roche Diagnostics. She is a member of the scientific advisory boards for Roche Diagnostics and Genentech and also consults for Diadem, DiamiR and Siemens Healthcare Diagnostics Inc. There are no conflicts.

Data availability

Proteomic data from the Knight ADRC participants are available at the NIAGADS and can be accessed at https://www.niagads.org/Knight ADRC-collection; Dataset NG00102: https://www.niagads.org/datasets/ng00102.

The summary results using these data are also available to the scientific community through a public web browser: http://ngi.pub:3838/ONTIME_Proteomics/. Data generated from the DIAN cohort can be requested at https://dian.wustl.edu/our-research/for-investigators/diantu-investigator-resources/dian-tu-biospecimen-request-form/.

Figures

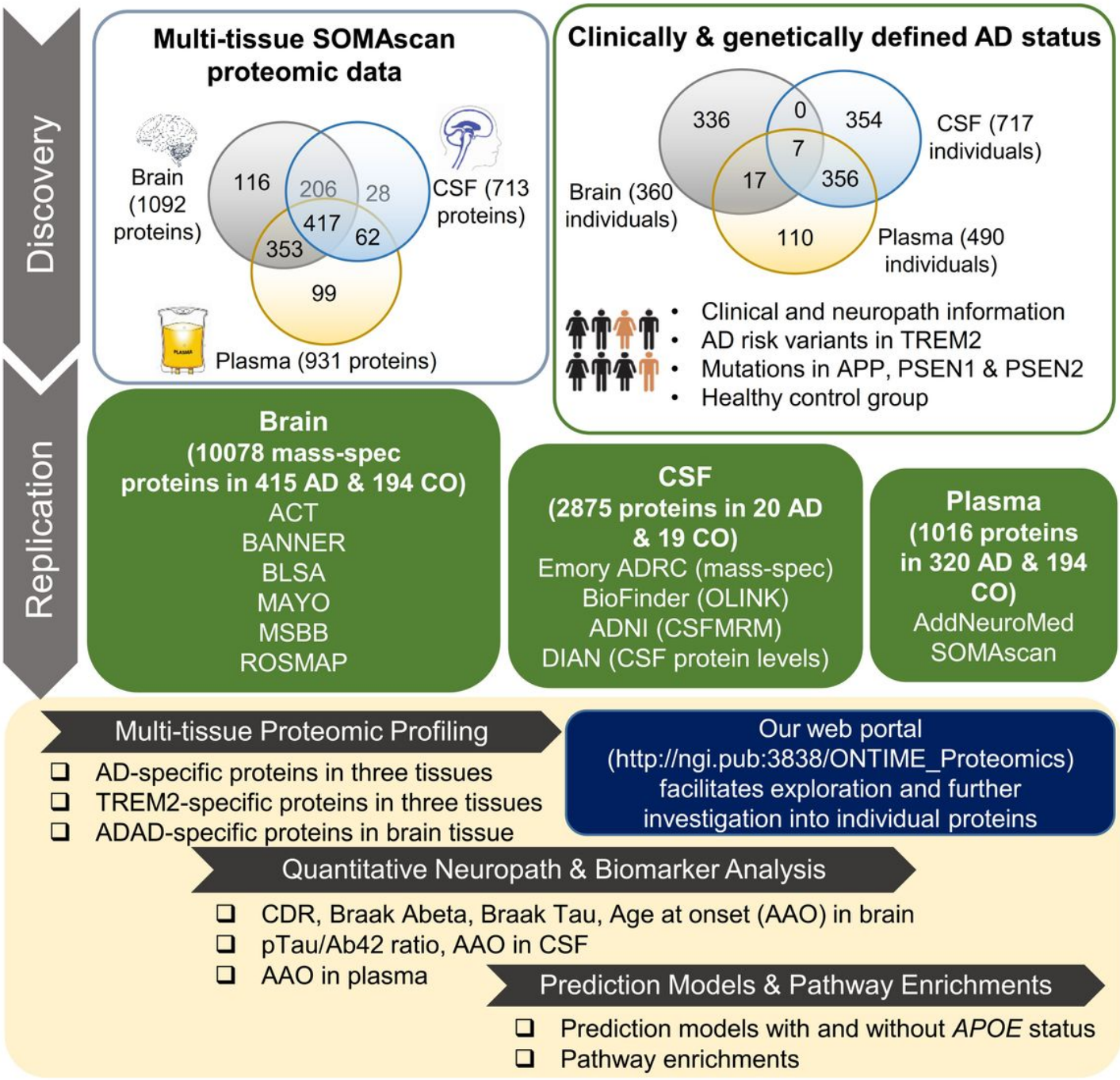

Figure 1

Study outline In discovery stage, protein measures with SOMAscan targeting 1,305 proteins were obtained in brain, CSF, and plasma tissues from well-characterized Knight ADRC and DIAN participants with comprehensive clinical information about AD pathology and cognition. This discovery cohort contained sporadic AD (290 in brain; 176 in CSF; 105 in plasma), TREM2 risk variant carriers (21 in brain; 47 in CSF; 131 in plasma), autosomal dominant AD (24 in brain), and healthy controls (25 in brain; 494 in CSF; 254 in plasma). Using this large number of samples, differential abundance analyses were

performed for sporadic AD status, TREM2 risk variant status and autosomal dominant AD status. Several publicly available external proteomics data were then used to replicate our findings. In addition, quantitative analyses using several quantitative neuropathology measures, including CDR and Braak scores in brain and age at onset in three tissues, were performed. Finally, replication proteins were used for creating a tissue-specific prediction models and pathway enrichment analysis. Our results are accessible through our web portal http://ngi.pub:3838/ONTIME_Proteomics/.

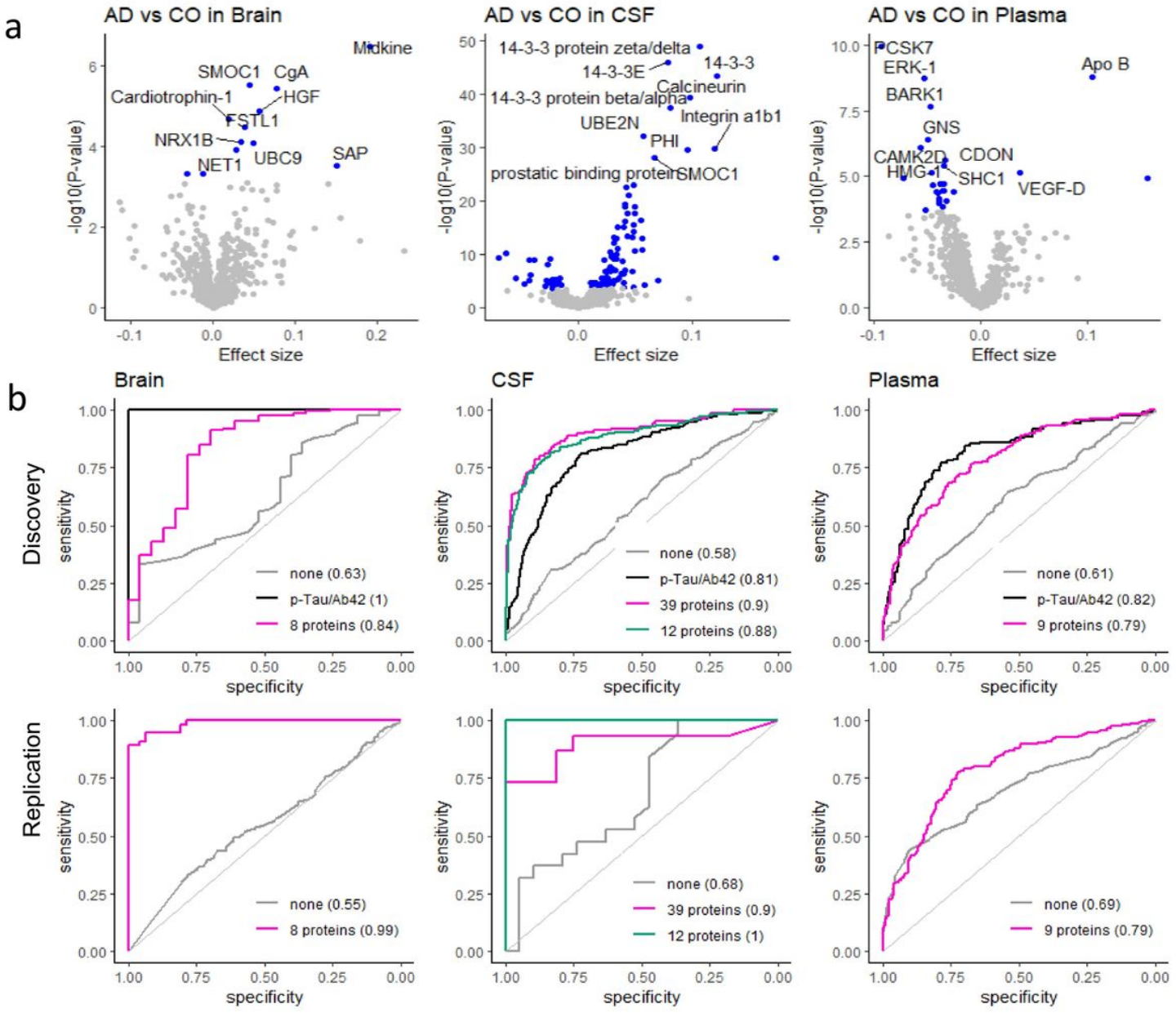

Figure 2

Multi-tissue proteomics profiling of sporadic AD a. Volcano plots are shown for brain, CSF, and plasma tissue. Multiple proteins (12 in brain; 117 in CSF; 26 in plasma; shown in blue) are differentially abundant in AD (compared to healthy controls) at the Bonferroni adjusted significance. A subset of those identified proteins also showed differential abundance levels in the other tissue. b. Tissue-specific prediction models are shown for our discovery data and externally replicated data set. For example, in CSF, among

117 proteins showing differential abundance levels in AD, 39 proteins (including SMOC1, Calcineurin, and ERK-1) were externally replicated showing nominal significance (P < 0.05) and same direction of effects. The prediction model using these 39 proteins provided an AUC of 0.89 in the Knight ADRC data and 0.9 in the Emory-ADRC data. In addition, the 12 proteins selected based on stepwise discriminant analysis (14-3-3 protein zeta/delta, EphA5, Calcineurin, Somatostatin-28, Cyclophilin A, Contactin-5, GFAP, Corticotropin-lipotropin, Spondin-1, TCTP, PolyUbiquitin K48, and Peroxiredoxin-6, Supplementary Table 14) led to an AUC of 0.88 in discovery and 0.999 in replication data. This showed that this predication model outperformed the gold standard CSF Aβ/tau181 ratio (with P = 2.4×10-6).

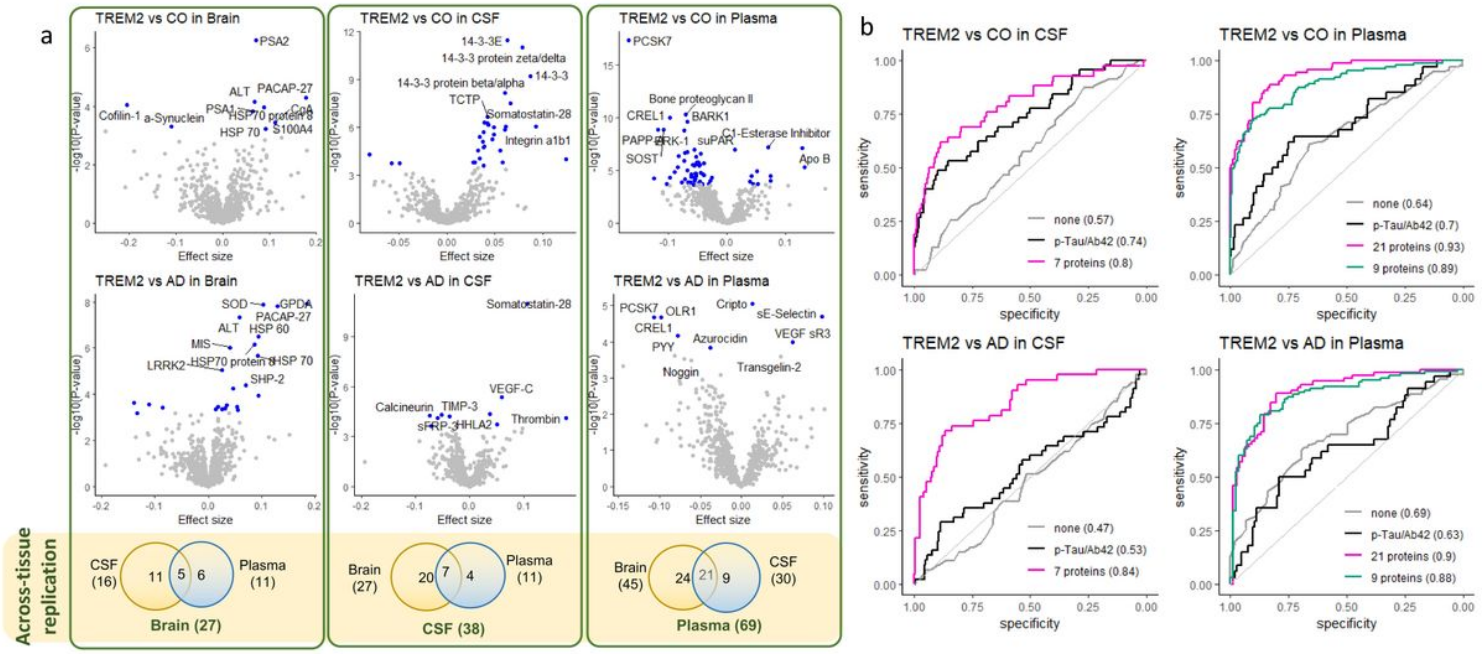

Figure 3

Multi-tissue proteomics profiling of TREM2 variant carrier status a. Volcano plots are shown for brain, CSF, and plasma tissue. Multiple proteins showed differential abundance levels in TREM2 variant carriers (compared to controls or other sporadic AD cases) in at least one of the three tissues. A subset of those identified proteins were replicated in the other tissues. For example, among 38 proteins associated with the TREM2 variant carrier status, 7 proteins (Supplementary Table 14) were replicated in brain and plasma. b. The prediction model based on the across-tissue replicated proteins (the pink curve with proteins validated by the other two tissues and the green curve based on the subset chosen through the discriminant analysis) showed higher accuracy than the well-accepted p-Tau/Aβ42 ratio (shown in black), while including age and sex as covariates. In plasma, out of 26 proteins, the 9 proteins selected based on stepwise discriminant analysis are Bone proteoglycan II, STAT3, uPA, ERK-1, VCAM-1, PAPP-A, BSSP4, XTP3A, and S100A4. The prediction models of our identified proteins while including age, sex and APOE status as covariates provided similar performance (Supplementary Fig. 6).

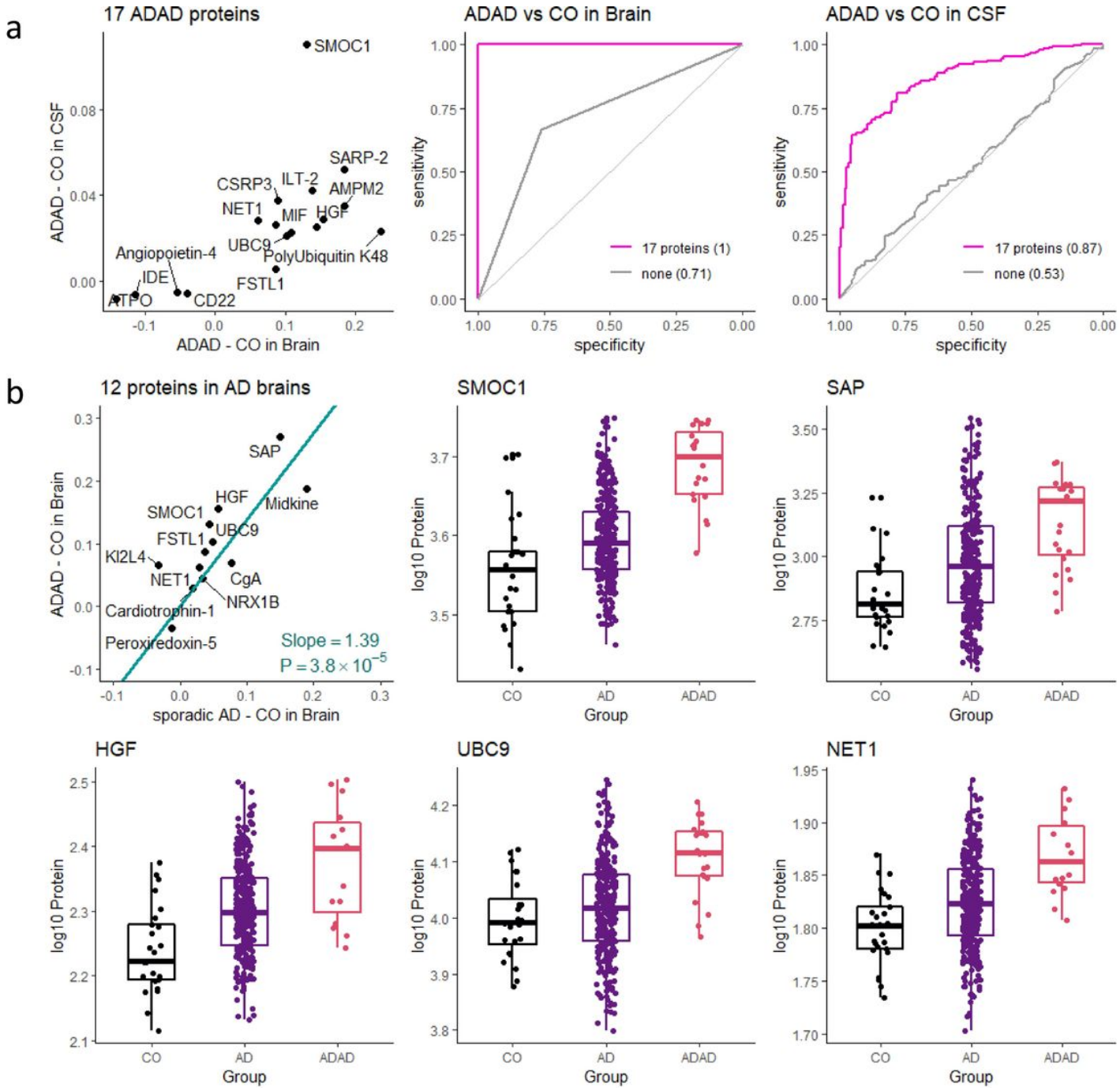

Figure 4

Proteomic profiling of autosomal dominant AD (ADAD) abundance. a. We found 109 proteins associated with the ADAD mutation carrier status at Bonferroni corrected threshold (volcano plot in Supplementary Fig. 5) and the 17 proteins (Supplementary Tables 12 and 14) were replicated in CSF and in the same direction. The model with these 17 proteins provided significantly higher AUC than the age alone (AUC = 1 vs 0.76; P = 9.9×10-3 in brain and AUC = 0.87 vs 0.53, P < 2.2×10-16 in CSF). b. The 12 proteins associated with sporadic AD brains displayed even stronger effect size in the ADAD mutation carrier brains. The effect of ADAD status on log-transformed protein levels (y-axis) roughly corresponded to 1.4 times the effect of AD status (x-axis) among the 12 identified proteins. This slope estimate (1.39, standard error = 0.21, P = 3.8×10-5) was obtained by fitting a regression line going through the origin,

which explained the scatterplot better than a regression line allowing the intercept (Multiple R-squared value = 0.80 without intercept vs. 0.65 with non-zero intercept). The box plots for the select 5 proteins are displayed.

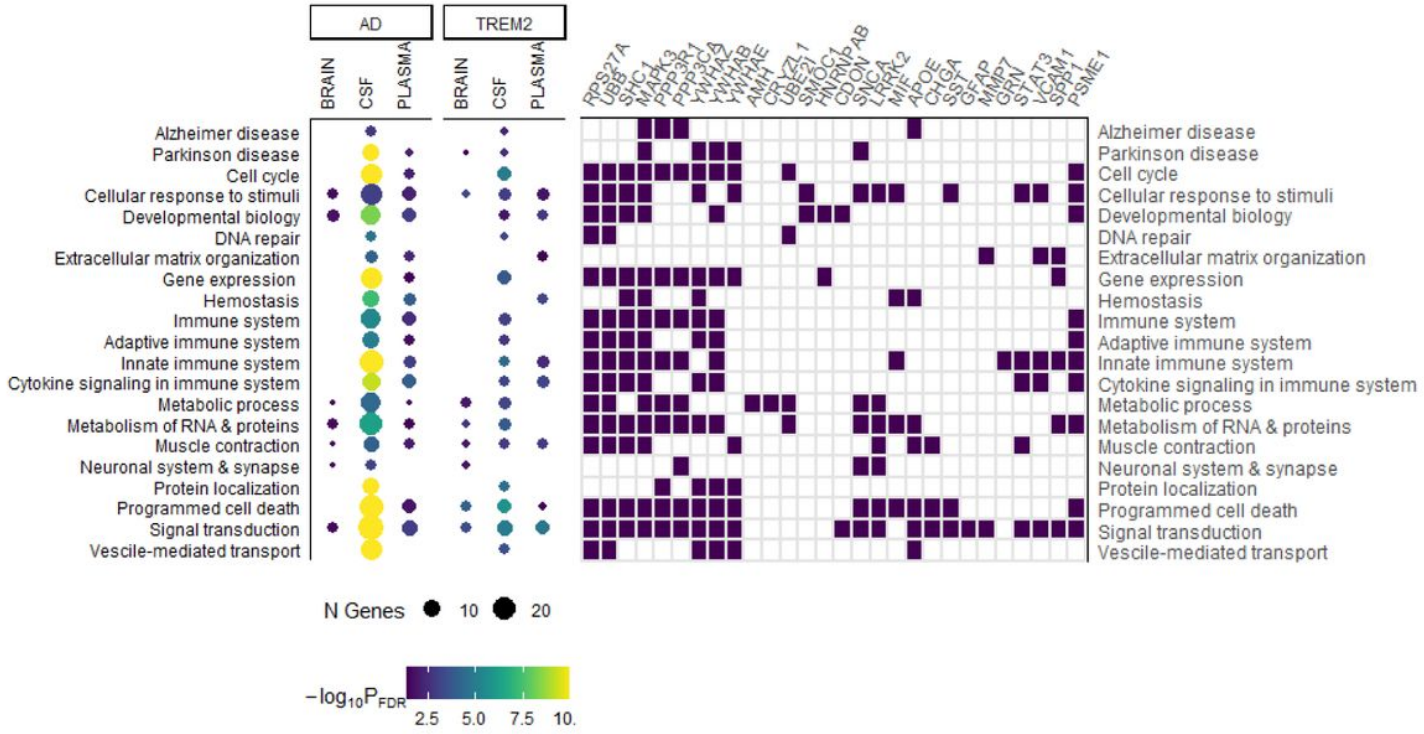

Figure 5

Pathway enrichments for multi-tissue findings The dot chart (size corresponding to the number of identified genes and color corresponding to the FDR corrected significance) presents that several identified proteins (Calcinuerin, APOE and α-synuclein) enrich in several pathways including Alzheimer disease, Parkinson disease and several immune related pathways (Supplementary Table 15). The full list of 79 genes is shown in Supplementary Fig. 8.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- MultitissueADproteomicsSupplTablesv4CCyj.xlsx
- MultitissueADproteomicsSuppFiguresv4CC.docx